


[Subscribe \(Full Service\)](#) [Register \(Limited Service, Free\)](#) [Login](#)

 Search: The ACM Digital Library The Guide

THE ACM DIGITAL LIBRARY
[Feedback](#)

(string and score and complete and address and record)

 Terms used: [string](#) [score](#) [complete](#) [address](#) [record](#)

 Sort results by relevance

Save results to a Binder

Refine these results

 Try this search in

 Display results expanded form

☐ Open results in a new window

Results 1 - 20 of 421

Result page: 1 2 3 4 5 6 7 8 9 10 next >>

1 [Matching records in a national medical patient index](#)



Glenn B. Bell, Anil Sethi

September 2001 Communications of the ACM, Volume 44 Issue 9

Publisher: ACM

 Full text available: pdf(146.20 KB) html(29.64 KB) Additional Information: [full citation](#), [references](#), [cited by](#), [index t](#)

Bibliometrics: Downloads (6 Weeks): 16, Downloads (12 Months): 152, Citation Count: 7

2 [Adaptive sorted neighborhood methods for efficient record linkage](#)



Su Yan, Dongwon Lee, Min-Yen Kan, Lee C. Giles

June 2007 JCDL '07: Proceedings of the 7th ACM/IEEE-CS joint conference on Digital libraries

Publisher: ACM

 Full text available: pdf(422.93 KB) Additional Information: [full citation](#), [abstract](#), [references](#), [index terms](#)

Bibliometrics: Downloads (6 Weeks): 11, Downloads (12 Months): 137, Citation Count: 0

Traditionally, record linkage algorithms have played an important role in maintaining digital libraries, identifying matching citations or authors for consolidation in updating or integrating digital libraries. As such, a variety of record linkage ...

 Keywords: [citation matching](#), [entity resolution](#), [record linkage](#), [sorted neighborhood](#)

3 [Automatic extraction of dynamic record sections from search engine result pages](#)

Hongkun Zhao, Weiyi Meng, Clement Yu

September 2006 VLDB '06: Proceedings of the 32nd international conference on Very large data ba

Publisher: VLDB Endowment

 Full text available: pdf(697.74 KB) Additional Information: [full citation](#), [abstract](#), [references](#), [index terms](#)

Bibliometrics: Downloads (6 Weeks): 15, Downloads (12 Months): 129, Citation Count: 0

A search engine returned result page may contain search results that are organized into multiple dynamically generated sections in response to a user query. Furthermore, such a result page often contains information irrelevant to the query, such ...

4 [Adaptive duplicate detection using learnable string similarity measures](#)



Mikhail Bilenko, Raymond J. Mooney

August 2003 KDD '03: Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining

Publisher: ACM

Full text available: [pdf\(239.92 KB\)](#)

Additional Information: full citation, abstract, references, cited by, index terms

Bibliometrics: Downloads (6 Weeks): 39, Downloads (12 Months): 231, Citation Count: 29

The problem of identifying approximately duplicate records in databases is an essential step for cleaning and data integration processes. Most existing approaches have relied on generic or manually tuned distance metrics for estimating the similarity ...

Keywords: SVM applications, data cleaning, distance metric learning, record linkage, string edit distance, trained similarity measures

5

Canonicalization of database records using adaptive similarity measures



Aron Culotta, Michael Wick, Robert Hall, Matthew Marzilli, Andrew McCallum

August 2007 KDD '07: Proceedings of the 13th ACM SIGKDD international conference on Knowledge discovery and data mining

Publisher: ACM

Full text available: [pdf\(852.22 KB\)](#)

Additional Information: full citation, abstract, references, index terms

Bibliometrics: Downloads (6 Weeks): 17, Downloads (12 Months): 214, Citation Count: 0

It is becoming increasingly common to construct databases from information automatically culled from many heterogeneous sources. For example, a research publication database can be constructed automatically extracting titles, authors, and conference ...

Keywords: data cleaning, data mining, information extraction

6

Canonicalization of graph database records using similarity measures



Na Li, Qing Li, Liping Wang

January 2008 IJCI MC '08: Proceedings of the 2nd international conference on Ubiquitous Information management and communication

Publisher: ACM

Full text available: [pdf\(232.94 KB\)](#)

Additional Information: full citation, abstract, references, index terms

Bibliometrics: Downloads (6 Weeks): 18, Downloads (12 Months): 32, Citation Count: 0

Information extraction and crawling from the Web have been increasingly common, yet raw data is often noisy and redundant due to heterogeneous sources. Although much work has focused on duplicate records detection, there is little investigation in providing ...

Keywords: canonicalization, database record, deduplication, graph mining

7

Collective entity resolution in relational data



Indrajit Bhattacharya, Lise Getoor

March 2007 ACM Transactions on Knowledge Discovery from Data (TKDD), Volume 1 Issue 1

Publisher: ACM

Full text available: [pdf\(511.57 KB\)](#)

Additional Information: full citation, abstract, references, index terms

Bibliometrics: Downloads (6 Weeks): 70, Downloads (12 Months): 701, Citation Count: 0

Many databases contain uncertain and imprecise references to real-world entities. The absence of unique identifiers for the underlying entities often results in a database which contains multiple references to the same entity. This can lead not only to data ...

Keywords: Entity resolution, data cleaning, graph clustering, record linkage

8 [A graph approach to the threshold all-against-all substring matching problem](#)



Marina Barsky, Ulrike Stege, Alex Thomo, Chris Upton

June 2008 Journal of Experimental Algorithms (JEA), Volume 12

Publisher: ACM

Full text available: pdf(615.01 KB)

Additional Information: [full citation](#), [abstract](#), [references](#), [index terms](#)

Bibliometrics: Downloads (6 Weeks): 3, Downloads (12 Months): 3, Citation Count: 0

We present a novel graph model and an efficient algorithm for solving the "threshold all against problem, which involves searching two strings (with length M and N , respectively) for all maxim approximate substring matches ...

Keywords: String matching, bioinformatics, complexity

9 [IEPAD: information extraction based on pattern discovery](#)



Chia-Hui Chang, Shao-Chen Lui

April 2001 WWW '01: Proceedings of the 10th international conference on World Wide Web

Publisher: ACM

Full text available: pdf(364.59 KB)

Additional Information: [full citation](#), [references](#), [cited by](#), [index terms](#)

Bibliometrics: Downloads (6 Weeks): 14, Downloads (12 Months): 167, Citation Count: 36

Keywords: PAT tree, extraction rule, information extraction, multiple string alignment

10 [ViPER: augmenting automatic information extraction with visual perceptions](#)



Kai Simon, Georg Lausen

October 2005 CIKM '05: Proceedings of the 14th ACM international conference on Information and knowledge management

Publisher: ACM

Full text available: pdf(484.44 KB)

Additional Information: [full citation](#), [abstract](#), [references](#), [cited by](#), [index terms](#)

Bibliometrics: Downloads (6 Weeks): 15, Downloads (12 Months): 84, Citation Count: 3

In this paper we address the problem of unsupervised Web data extraction. We show that unsupervised Web data extraction becomes feasible when supposing pages that are made up of repetitive patterns as it is the case, e.g., for search engine result pages. ...

Keywords: data extraction, data record alignment, visual features

11 [Querying web metadata: Native score management and text support in databases](#)



Gültekin Özsoyoğlu, İsmail Sengör Altıngöve, Abdullah Al-Hamdani, Selma Ayşe Özel, Özgür Ulus, Zehra Meral Özsoyoğlu

December 2004 ACM Transactions on Database Systems (TODS), Volume 29 Issue 4

Publisher: ACM

Full text available: pdf(737.76 KB)

Additional Information: [full citation](#), [appendices and supplements](#), [abstract](#), [references](#), [cited by](#), [index terms](#)

Bibliometrics: Downloads (6 Weeks): 25, Downloads (12 Months): 169, Citation Count: 1

In this article, we discuss the issues involved in adding a native score management system to relational databases, to be used in querying Web metadata (that describes the semantic content of resources). The Web metadata model is based on ...

Keywords: Score management for Web applications

12 [Febri: a freely available record linkage system with a graphical user interface](#)

Peter Christen

January 2008 HDKM '08: Proceedings of the second Australasian workshop on Health data knowledge management - Volume 80, Volume 80

Publisher: Australian Computer Society, Inc.

Full text available:  pdf(320.97 KB)

Additional Information: full citation, abstract, references

Bibliometrics: Downloads (6 Weeks): 0, Downloads (12 Months): 0, Citation Count: 0

Record or data linkage is an important enabling technology in the health sector, as linked data is a cost-effective resource that can help to improve research into health policies, detect adverse drug reactions, reduce costs, and uncover fraud within ...


Keywords: GUI, data cleaning, data integration, data matching, deduplication, health data linkage, open source software, record linkage software

13 [One-dimensional and multi-dimensional substring selectivity estimation](#)

H. V. Jagadish, Olga Kapitskaia, Raymond T. Ng, Divesh Srivastava

December 2000 The VLDB Journal — The International Journal on Very Large Data Bases, Volume 9, Issue 3

Publisher: Springer-Verlag New York, Inc.

Full text available:  pdf(215.66 KB)

Additional Information: full citation, abstract, references, cited by, index term

Bibliometrics: Downloads (6 Weeks): 4, Downloads (12 Months): 18, Citation Count: 4

With the increasing importance of XML, LDAP directories, and text-based information sources or Internet, there is an ever-greater need to evaluate queries involving (sub)string matching. In many cases, matches need to be on multiple attributes/dimensions, ...


Keywords: Maximal overlap, Pruned count-suffix tree, Short memory property, String selectivity

14 [Web application security assessment by fault injection and behavior monitoring](#)

Yao-Wen Huang, Shih-Kun Huang, Tsung-Po Lin, Chung-Hung Tsai

May 2003 WWW '03: Proceedings of the 12th international conference on World Wide Web

Publisher: ACM

Full text available:  pdf(4.63 MB)

Additional Information: full citation, abstract, references, cited by, index term

Bibliometrics: Downloads (6 Weeks): 105, Downloads (12 Months): 689, Citation Count: 11

As a large and complex application platform, the World Wide Web is capable of delivering a broad range of sophisticated applications. However, many Web applications go through rapid development cycles with extremely short turnaround time, making it difficult ...

Keywords: black-box testing, complete crawling, fault injection, security assessment, web application testing

15 [A cost-based model and effective heuristic for repairing constraints by value modification](#)

Philip Bohannon, Wenfei Fan, Michael Flaster, Rajeev Rastogi

June 2005 SIGMOD '05: Proceedings of the 2005 ACM SIGMOD international conference on Management of data

Publisher: ACM

Full text available:  pdf(565.83 KB)

Additional Information: full citation, abstract, references, cited by

Bibliometrics: Downloads (6 Weeks): 12, Downloads (12 Months): 84, Citation Count: 3

Data integrated from multiple sources may contain inconsistencies that violate integrity constraints. The *constraint repair problem* attempts to find "low cost" changes that, when applied, will cause constraints to be satisfied. While in most ...


16 Typing Staphylococcus aureus Using the spa Gene and Novel Distance Measures

Phaedra Agius, Barry Kreiswirth, Steve Naidich, Kristin Bennett

October 2007 IEEE/ACM Transactions on Computational Biology and Bioinformatics (TCBB)

Volume 4 Issue 4

Publisher: IEEE Computer Society Press

Full text available:  pdf(2.81 MB)

Additional Information: [full citation](#), [abstract](#), [index terms](#)

Bibliometrics: Downloads (6 Weeks): 6, Downloads (12 Months): 48, Citation Count: 0

We developed an approach for identifying groups or families of Staphylococcus aureus bacteria I on genotype data. With the emergence of drug resistant strains, S. aureus represents a significant human health threat. Identifying the family types efficiently ...

Keywords: clustering, sequence algorithms, genotyping, staphylococcus aureus, molecular epidemiology


17 An embedded domain-specific language for type-safe server-side web scripting



Peter Thiemann

February 2005 ACM Transactions on Internet Technology (TOIT), Volume 5 Issue 1

Publisher: ACM

Full text available:  pdf(336.60 KB)

Additional Information: [full citation](#), [abstract](#), [references](#), [cited by](#), [index term review](#)

Bibliometrics: Downloads (6 Weeks): 138, Downloads (12 Months): 347, Citation Count: 2

WASH/CGI is an embedded domain-specific language for server-side Web scripting. Due to its reliance on the strongly typed, purely functional programming language Haskell as a host language, it is flexible and---at the same time---it provides extensive ...

Keywords: Interactive Web services, Web programming

18 Creation of an expert witness database through text mining



Christopher Dozier, Peter Jackson, Xi Guo, Mark Chaudhary, Yohendran Arumainayagam

June 2003 ICAIL '03: Proceedings of the 9th international conference on Artificial Intelligence and

Publisher: ACM

Full text available:  pdf(244.79 KB)

Additional Information: [full citation](#), [abstract](#), [references](#), [cited by](#)

Bibliometrics: Downloads (6 Weeks): 14, Downloads (12 Months): 76, Citation Count: 1

This paper describes how an online directory of expert witnesses was created from jury verdict settlement documents using text mining techniques. We have created an expert witness directory contains over 100,000 expert profiles, based on approximately ...

Keywords: expert witness, information extraction, text mining


19 Trustworthy 100-year digital objects: Evidence after every witness is dead



Henry M. Gladney

July 2004 ACM Transactions on Information Systems (TOIS), Volume 22 Issue 3

Publisher: ACM

Full text available:  pdf(1.24 MB)

Additional Information: full citation, abstract, references, cited by, index terms

Bibliometrics: Downloads (6 Weeks): 32, Downloads (12 Months): 205, Citation Count: 4

In ancient times, wax seals impressed with signet rings were affixed to documents as evidence of authenticity. A digital counterpart is a message authentication code fixed firmly to each important document. If a digital object is sealed together ...


20 Articulating information needs in XML query languages



Jaap Kamps, Maarten Marx, Maarten de Rijke, Börkur Sigurbjörnsson

October 2006 ACM Transactions on Information Systems (TOIS), Volume 24 Issue 4

Publisher: ACM

Full text available:  pdf(318.47 KB)

Additional Information: full citation, abstract, references, index terms

Bibliometrics: Downloads (6 Weeks): 14, Downloads (12 Months): 159, Citation Count: 0

Document-centric XML is a mixture of text and structure. With the increased availability of document-centric XML documents comes a need for query facilities in which both structural constraints and constraints on the content of the documents can be expressed. ...

Keywords: Full-text XML querying, XML retrieval, XPath

Results 1 - 20 of 421

Result page: 1 2 3 4 5 6 7 8 9 10 next >>

The ACM Portal is published by the Association for Computing Machinery. Copyright © 2008 ACM

[Terms of Usage](#) [Privacy Policy](#) [Code of Ethics](#) [Contact Us](#)

Useful downloads:  Adobe Acrobat  QuickTime  Windows Media Player  RealPlayer